K-Means Clustering of People with COVID-19

July 2, 2020

DoYeong, Jeon
Software Engineering Laboratory
Soongsil University

TABLE OF CONTENTS

1. Source Code

1.1. Code for Creating Database

❑ CreatingDB Class

```
class CreatingDB:
    """
    Class for creating random database
    """
    num_people = 0  # number of people to create
    base_date = None  # the base date of data

    def __init__(self, num_people, base_date):
        self.num_people = num_people
        self.base_date = base_date

    def generate_incurred_date(self):
        """
        function to create random incurred date
        :return:
            incurred_date: string, the day of infection or contact
            elapsed_days: int, the difference between base date and incurred
date
        """
        elapsed_days = random.randint(0, 14)  # the valid day period is 0~14
        # extracting the incurred day using periods and base date
        incurred_date = (self.base_date - timedelta(days=elapsed_days)). \
            strftime("%Y %m %d")
        return incurred_date, elapsed_days

    def generate_address_list(self):
        """
        function to get one address randomly from the adress list
        :return: the randomly generated address list
        """
        with open('./Address_Part.txt', 'r', encoding='utf-8') as add_file:
            # add_file = add_file.encoding
            address_list = add_file.readlines()

            random_address_list = []  # list to store addresses

            # extract addresses as many as the number of recipients
            for _ in range(1, self.num_people + 1):
                random_address_list.append(random.choice(address_list))

        return random_address_list

    def generate_csv_data(self):
        """
        function to create .csv file with randomly generated records
        :return: None
        """
        num_healthy = round(self.num_people / 3)  # 1/3 is healthy
        num_contacted = round(self.num_people / 3)  # 1/3 is contacted
        # 1/3 is confirmed
        num_confirmed = self.num_people - num_healthy - num_contacted

        id_list = list(range(1, self.num_people + 1))  # ID as many as people
        random.shuffle(id_list)  # shuffle list

        # age records as many as people
        age_list = list(random.randint(1, 100)
                       for _ in range(1, self.num_people + 1))
        # address records as many as people
        address_list = self.generate_address_list()

        severity_list = []  # severity records as many as people
        incurred_date_list = []  # incurred date list including 'None'(healthy)
        status_list = []  # status(Healthy, Contacted, and Confirmed) list

        # Entire people num = healthy + contacted + confirmed
        # Repeat as many healthy people
        for _ in range(num_healthy):
            # severity_list.append(0)
            status_list.append('Healthy')
            incurred_date_list.append('None')

        # Repeat as many contacted people
        for count in range(num_contacted):
            date, days = self.generate_incurred_date()
            status_list.append('Contacted')
            #    severity_list.append(round(self.compute_severity('contacted',
days), 2))
            incurred_date_list.append(date)

        # Repeat as many confirmed people
        for _ in range(num_confirmed):
            date, days = self.generate_incurred_date()
            status_list.append('Confirmed')
            #    severity_list.append(round(self.compute_severity('confirmed',
days), 2))
            incurred_date_list.append(date)

        # converting as pandas DataFrame data type to save .csv
        df = pd.DataFrame({
            "ID": id_list,
            "Age": age_list,
            "Address": address_list,
            "Covid Status": status_list,
            # "Severity": severity_list,
            "Incurred Date": incurred_date_list,
        })
        df = df.sort_values(['ID'], ascending=[True])
        df.reset_index(drop=True, inplace=True)

        # saving as .csv file
        df.to_csv("corona_data.csv", mode='w', encoding='utf-8-sig')
```

1.2. Code for Clustering

❑ **ClusteringPeople Class**

```
from datetime import datetime, timedelta
from dateutil.parser import parse
import random
import pandas as pd
import numpy as np
import sklearn
from sklearn import cluster
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt
from matplotlib import cm


class CreatingDB:
    """
    Class for creating random database
    """
    num_people = 0  # number of people to create
    base_date = None  # the base date of data

    def __init__(self, num_people, base_date):
        self.num_people = num_people
        self.base_date = base_date

    def generate_incurred_date(self):
        """
        function to create random incurred date
        :return:
            incurred_date: string, the day of infection or contact
            elapsed_days: int, the difference between base date and incurred
date
        """
        elapsed_days = random.randint(0, 14)  # the valid day period is 0~14
        # extracting the incurred day using periods and base date
        incurred_date = (self.base_date - timedelta(days=elapsed_days)). \
            strftime("%Y %m %d")
        return incurred_date, elapsed_days

    def generate_address_list(self):
        """
        function to get one address randomly from the adress list
        :return: the randomly generated address list
        """
        with open('./Address_Part.txt', 'r', encoding='utf-8') as add_file:
            # add_file = add_file.encoding
            address_list = add_file.readlines()
            random_address_list = []  # list to store addresses

            # extract addresses as many as the number of recipients
            for _ in range(1, self.num_people + 1):
                random_address_list.append(random.choice(address_list))

        return random_address_list

    def generate_csv_data(self):
        """
        function to create .csv file with randomly generated records
        :return: None
        """
        num_healthy = round(self.num_people / 3)  # 1/3 is healthy
        num_contacted = round(self.num_people / 3)  # 1/3 is contacted
        # 1/3 is confirmed
        num_confirmed = self.num_people - num_healthy - num_contacted

        id_list = list(range(1, self.num_people + 1))  # ID as many as people
        random.shuffle(id_list)  # shuffle list

        # age records as many as people
        age_list = list(random.randint(1, 100)
                       for _ in range(1, self.num_people + 1))
        # address records as many as people
        address_list = self.generate_address_list()

        severity_list = []  # severity records as many as people
        incurred_date_list = []  # incurred date list including 'None'(healthy)
        status_list = []  # status(Healthy, Contacted, and Confirmed) list

        # Entire people num = healthy + contacted + confirmed
        # Repeat as many healthy people
        for _ in range(num_healthy):
            # severity_list.append(0)
            status_list.append('Healthy')
            incurred_date_list.append('None')

        # Repeat as many contacted people
        for count in range(num_contacted):
            date, days = self.generate_incurred_date()
            status_list.append('Contacted')
            #     severity_list.append(round(self.compute_severity('contacted',
days), 2))
            incurred_date_list.append(date)

        # Repeat as many confirmed people
        for _ in range(num_confirmed):
            date, days = self.generate_incurred_date()
            status_list.append('Confirmed')
            #     severity_list.append(round(self.compute_severity('confirmed',
days), 2))
            incurred_date_list.append(date)
        # converting as pandas DataFrame data type to save .csv
        df = pd.DataFrame({
            "ID": id_list,
            "Age": age_list,
            "Address": address_list,
            "Covid Status": status_list,
            # "Severity": severity_list,
            "Incurred Date": incurred_date_list,
        })
        df = df.sort_values(['ID'], ascending=[True])
        df.reset_index(drop=True, inplace=True)

        # saving as .csv file
        df.to_csv("corona_data.csv", mode='w', encoding='utf-8-sig')


class ClusteringPeople:
    df_corona = None
    cluster_result_dic = {}

    def __init__(self, file_path):
        self.load_data(file_path)

    def load_data(self, file_path):
        """
        method to load .csv file
        :param file_path: string, the path of file
        :return:
        """
        self.df_corona = pd.read_csv(file_path)

    def display_load_data(self):
        print(f"{'ID':<4}{'Age':<4}{'Covid
Status':<13}{'Severity':<9}{'Address':<10}")
        for i in range(len(self.df_corona)):
            print(f"{self.df_corona['ID'][i]:<4}"
                  f"{self.df_corona['Age'][i]:<4}"
                  f"{self.df_corona['Covid Status'][i]:<13}"
                  f"{round(self.df_corona['Severity'][i], 3):<9}"
                  f"{self.df_corona['Address'][i].split()[0]:<10}"
                  )
        print()  # float 1 line

    def preprocess(self):
        """
        method to preprocess the data for distance function
        :return: None
        """
        col_num = len(self.df_corona)  # the number of rows from loaded data
        today = datetime.now().date()  # date of today, YEAR-MONTH-DAY

        # selecting specific column to compute 'severity'
        incur_date_col = self.df_corona['Incurred Date']
        status = self.df_corona['Covid Status']
        severity_list = []  # list for storing severity result

        for i in range(col_num):
            severity = 0  # default is healthy, 0.
            if status[i] == 'Contacted':  # contacted person?
                # formula for contacted person:
                #   x = 1 - ((today's date) - (infected date)) * 0.05)
                elapsed_days = (today - parse(incur_date_col[i]).date()).days
                severity = 1 - (elapsed_days * 0.05)
            elif status[i] == 'Confirmed':  # confirmed person?
                # formula for confirmed person:
                #   x = (1 - ((today's date) - (infected date)) * 0.05)) / 2
                elapsed_days = (today - parse(incur_date_col[i]).date()).days
                severity = (1 - (elapsed_days * 0.05)) * 0.5

            severity_list.append(severity)  # add the value to the list
        self.df_corona["Severity"] = severity_list

    def cluster(self):
        sse_list = []  # list for storing SSE(Sum of squares errors)
        silhouette_score_list = []  # list for storing silhouette scores

        for i in range(2, 10):  # number of clusters 2 to 9
            # load the k-means model
            km = cluster.KMeans(
                n_clusters=i,  # the number of cluster
                init='k-means++',  # how to initial cluster centers
                max_iter=300,  # maximum number of iterations
                algorithm='auto'  # three choices: auto, full, and elkan.
            )

            # changing the shape of data
            severity_list = self.df_corona["Severity"].values.tolist()
            severity_list = np.array(severity_list)

            # cluster
            cluster_predicted_list  =  km.fit_predict(severity_list.reshape(-1,
1))

            # storing SSE value to get the optimal number of cluster
            sse_list.append(km.inertia_)

            # storing silhouette score to get optimal number of cluster

silhouette_score_list.append(silhouette_score(severity_list.reshape(-1,   1),
cluster_predicted_list))

            cluster_list = [j for j in range(i)]  # cluster list
            # display the reuslt of cluster
            self.print_result_of_cluster(cluster_list, cluster_predicted_list)

            # store the prediction result
            self.cluster_result_dic[i] = cluster_predicted_list
    def draw_elbow_method(self, sse_list):
        """
        method to draw elbow graph using SSE(Sum of Squares Error)
        :param sse_list: list of SSE
        :return: None
        """
        plt.plot(range(2, 10), sse_list, marker='o')
        plt.xlabel("The Number of Cluster")
        plt.ylabel("SSE")
        plt.show()

    def print_result_of_cluster(self, cluster_list, cluster_predicted_list):
        """
        form
        Cluster 1:
            Number of people: n
            Severity Values: [...]
            Average of severities: n
        Cluster 2:
            ...

        :return:
        """
        severity_list = self.df_corona["Severity"].values.tolist()
        id_list = self.df_corona["ID"].values.tolist()

        cluster_predicted_list = cluster_predicted_list.tolist()
        print(f"The number of Cluster: {len(cluster_list)}")
        for cluster_idx in cluster_list:
            num_people = cluster_predicted_list.count(cluster_idx)
            id_severity_tuple_list = []
            sum_of_severities = 0
            for person_idx in range(len(cluster_predicted_list)):
                if cluster_idx == cluster_predicted_list[person_idx]:
                    sum_of_severities += severity_list[person_idx]
                    id_severity_tuple_list.append((person_idx+1,
round(severity_list[person_idx], 2)))
            print(f"\tCluster {cluster_idx}:")
            print(f"\t\tNumber of People: {num_people}")
            print(f"\t\tPeople list with Severity Values:")
            print(f"\t\t\t{'ID':<4}{'Severity Value'}")
            for person_in_cluster in id_severity_tuple_list:
                if id_severity_tuple_list.index(person_in_cluster) % 2 == 0:

print(f"\t\t\t{person_in_cluster[0]:<4}{person_in_cluster[1]}")

            print(f"\t\tAverage of severities: {round(sum_of_severities /
len(id_severity_tuple_list), 2)}")
        print()  # float 1 line


    def draw_silhouette(self):
        """
        method to draw graph using silhouette scores
        :return: None
        """
        pass

    def draw_graph(self):
        """
        method to draw clustering result
        :return: None
        """
        pass


if __name__ == '__main__':
    # CODE FOR CREATING DATABASE
    # require the number of people and base date
    # num_people = int(input("Enter the number of people: "))
    # date_input = input("Enter the base date(Year-Month-Day): ")
    # if date_input == '':
    #     print("The base date is set as today.")
    #     date = datetime.now().date()
    # else:
    #     date = parse(date_input).date()
    #
    # cdb = CreatingDB(num_people, date)  # creating instance
    # cdb.generate_csv_data()  # creating .csv file

    # CODE FOR CLUSTERING
    file_path = './corona_data.csv'

    cp = ClusteringPeople(file_path)
    cp.preprocess()
    cp.display_load_data()
    cp.cluster()
    # cp.draw_graph()
```

- **main**

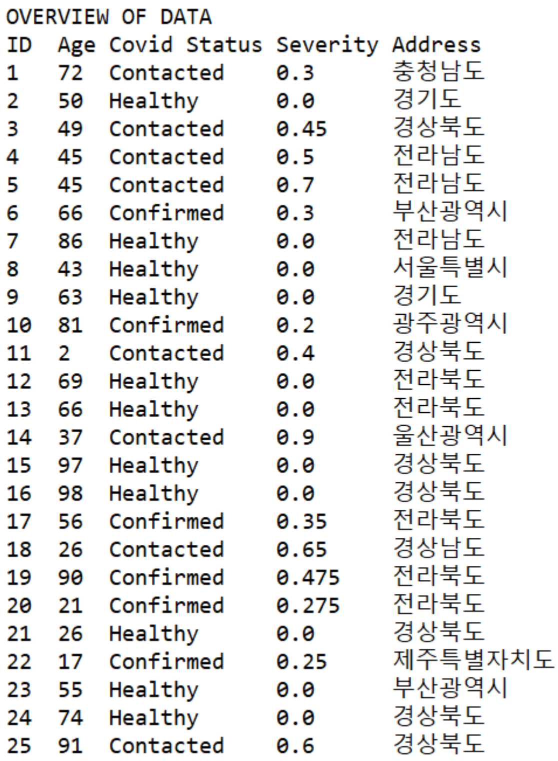
```
if __name__ == '__main__':
    # CODE FOR CLUSTERING
    file_path = './corona_data.csv'

    cp = ClusteringPeople(file_path)
    cp.preprocess()
    cp.draw_graph()
    cp.cluster()
```

2. Result of Clustering

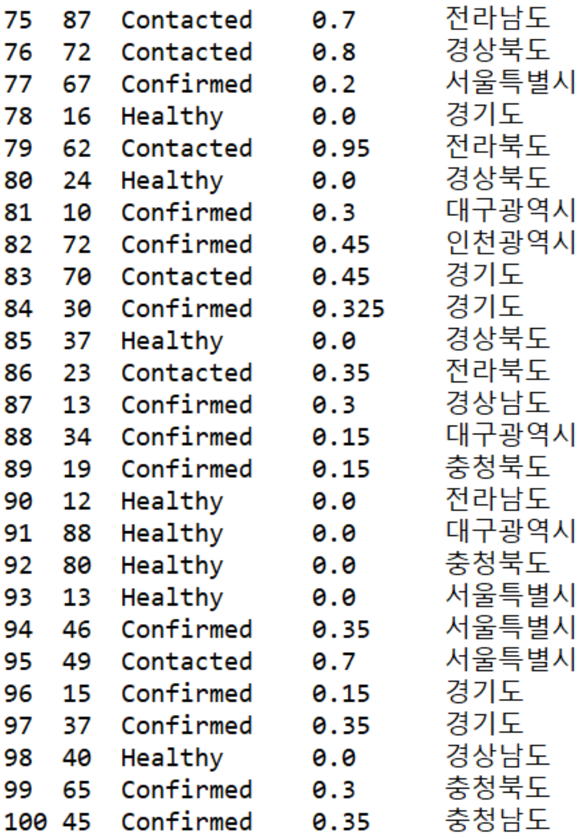
2.1. Loaded Dataset

- **Top 25 lines**



```
OVERVIEW OF DATA
ID  Age Covid Status Severity Address
1   72  Contacted    0.3      충청남도
2   50  Healthy      0.0      경기도
3   49  Contacted    0.45     경상북도
4   45  Contacted    0.5      전라남도
5   45  Contacted    0.7      전라남도
6   66  Confirmed    0.3      부산광역시
7   86  Healthy      0.0      전라남도
8   43  Healthy      0.0      서울특별시
9   63  Healthy      0.0      경기도
10  81  Confirmed    0.2      광주광역시
11  2   Contacted    0.4      경상북도
12  69  Healthy      0.0      전라북도
13  66  Healthy      0.0      전라북도
14  37  Contacted    0.9      울산광역시
15  97  Healthy      0.0      경상북도
16  98  Healthy      0.0      경상북도
17  56  Confirmed    0.35     전라북도
18  26  Contacted    0.65     경상남도
19  90  Confirmed    0.475    전라북도
20  21  Confirmed    0.275    전라북도
21  26  Healthy      0.0      경상북도
22  17  Confirmed    0.25     제주특별자치도
23  55  Healthy      0.0      부산광역시
24  74  Healthy      0.0      경상북도
25  91  Contacted    0.6      경상북도
```

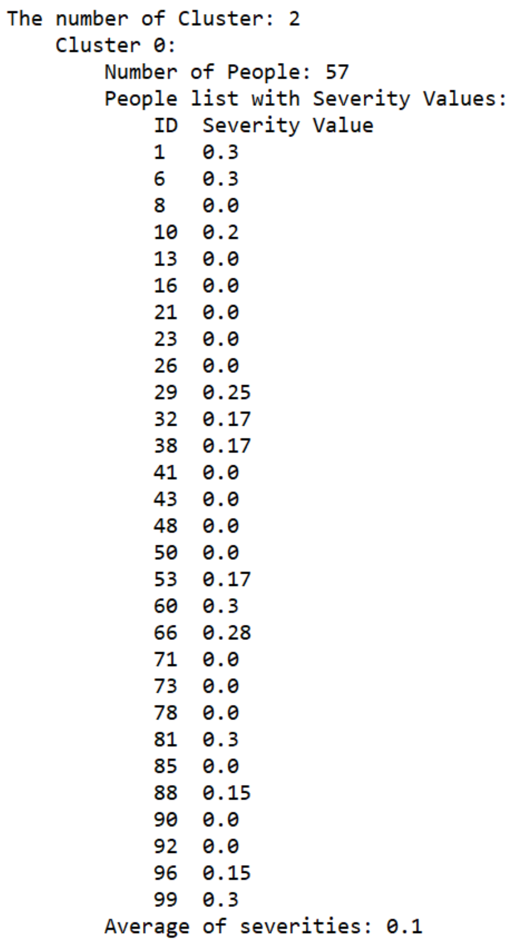
- **Last 25 lines**



```
75  87  Contacted    0.7      전라남도
76  72  Contacted    0.8      경상북도
77  67  Confirmed    0.2      서울특별시
78  16  Healthy      0.0      경기도
79  62  Contacted    0.95     전라북도
80  24  Healthy      0.0      경상북도
81  10  Confirmed    0.3      대구광역시
82  72  Confirmed    0.45     인천광역시
83  70  Contacted    0.45     경기도
84  30  Confirmed    0.325    경기도
85  37  Healthy      0.0      경상북도
86  23  Contacted    0.35     전라북도
87  13  Confirmed    0.3      경상남도
88  34  Confirmed    0.15     대구광역시
89  19  Confirmed    0.15     충청북도
90  12  Healthy      0.0      전라남도
91  88  Healthy      0.0      대구광역시
92  80  Healthy      0.0      충청북도
93  13  Healthy      0.0      서울특별시
94  46  Confirmed    0.35     서울특별시
95  49  Contacted    0.7      서울특별시
96  15  Confirmed    0.15     경기도
97  37  Confirmed    0.35     경기도
98  40  Healthy      0.0      경상남도
99  65  Confirmed    0.3      충청북도
100 45  Confirmed    0.35     충청남도
```

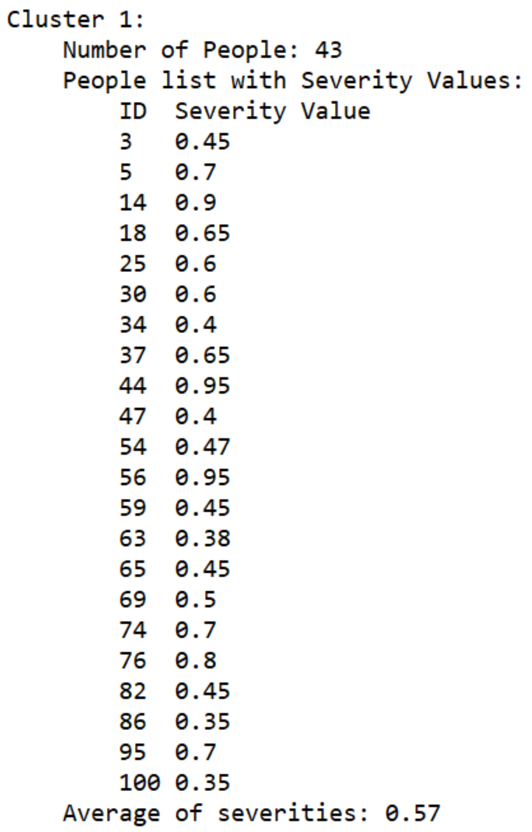
2.2. K-Means

- **The number of Cluster: 2**
  - Cluster 0



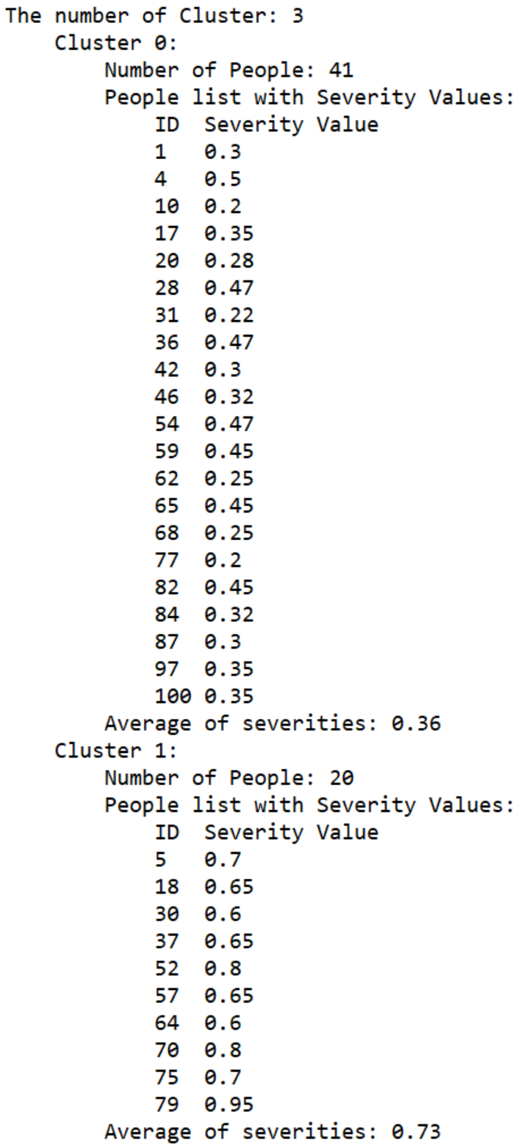
```
The number of Cluster: 2
    Cluster 0:
        Number of People: 57
        People list with Severity Values:
            ID  Severity Value
            1   0.3
            6   0.3
            8   0.0
            10  0.2
            13  0.0
            16  0.0
            21  0.0
            23  0.0
            26  0.0
            29  0.25
            32  0.17
            38  0.17
            41  0.0
            43  0.0
            48  0.0
            50  0.0
            53  0.17
            60  0.3
            66  0.28
            71  0.0
            73  0.0
            78  0.0
            81  0.3
            85  0.0
            88  0.15
            90  0.0
            92  0.0
            96  0.15
            99  0.3
        Average of severities: 0.1
```

- Cluster 1



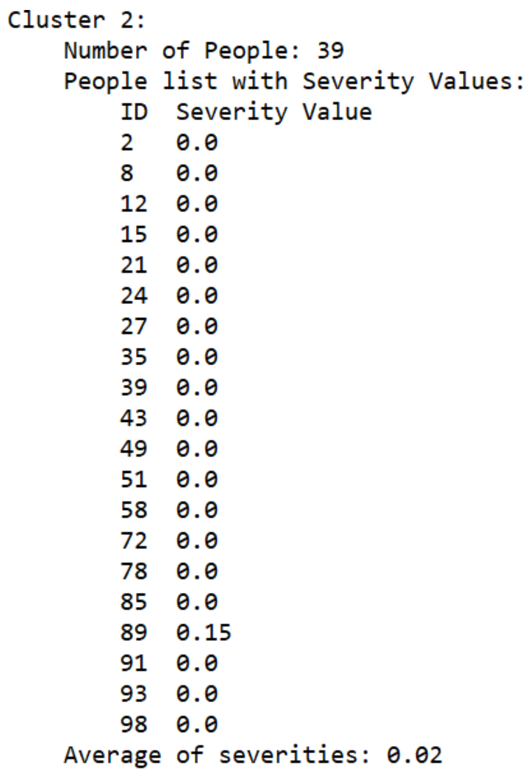
```
Cluster 1:
    Number of People: 43
    People list with Severity Values:
        ID  Severity Value
        3   0.45
        5   0.7
        14  0.9
        18  0.65
        25  0.6
        30  0.6
        34  0.4
        37  0.65
        44  0.95
        47  0.4
        54  0.47
        56  0.95
        59  0.45
        63  0.38
        65  0.45
        69  0.5
        74  0.7
        76  0.8
        82  0.45
        86  0.35
        95  0.7
        100 0.35
    Average of severities: 0.57
```

- **The number of Cluster: 3**
  - Cluster 0, 1



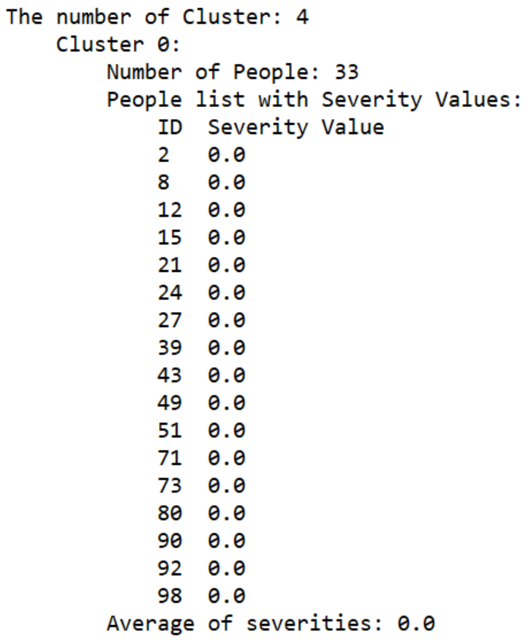
```
The number of Cluster: 3
    Cluster 0:
        Number of People: 41
        People list with Severity Values:
            ID  Severity Value
            1   0.3
            4   0.5
            10  0.2
            17  0.35
            20  0.28
            28  0.47
            31  0.22
            36  0.47
            42  0.3
            46  0.32
            54  0.47
            59  0.45
            62  0.25
            65  0.45
            68  0.25
            77  0.2
            82  0.45
            84  0.32
            87  0.3
            97  0.35
            100 0.35
        Average of severities: 0.36
    Cluster 1:
        Number of People: 20
        People list with Severity Values:
            ID  Severity Value
            5   0.7
            18  0.65
            30  0.6
            37  0.65
            52  0.8
            57  0.65
            64  0.6
            70  0.8
            75  0.7
            79  0.95
        Average of severities: 0.73
```

  - Cluster 2



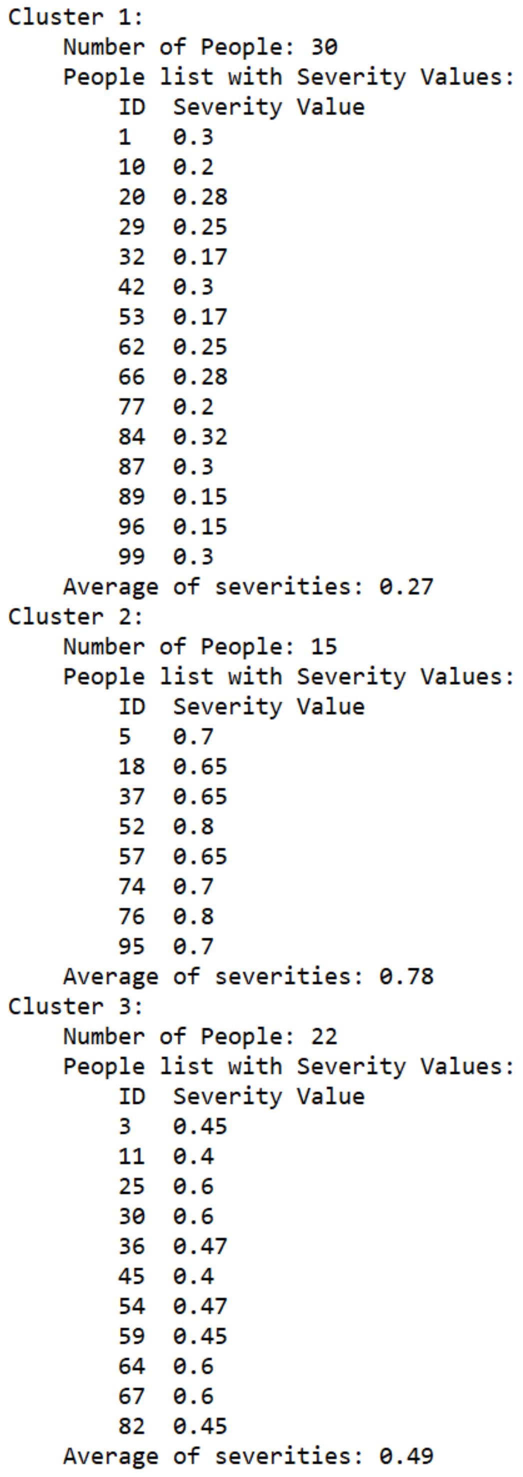
```
Cluster 2:
    Number of People: 39
    People list with Severity Values:
        ID  Severity Value
        2   0.0
        8   0.0
        12  0.0
        15  0.0
        21  0.0
        24  0.0
        27  0.0
        35  0.0
        39  0.0
        43  0.0
        49  0.0
        51  0.0
        58  0.0
        72  0.0
        78  0.0
        85  0.0
        89  0.15
        91  0.0
        93  0.0
        98  0.0
    Average of severities: 0.02
```

- ❑ **The number of Cluster: 4**
  - ❍ Cluster 0



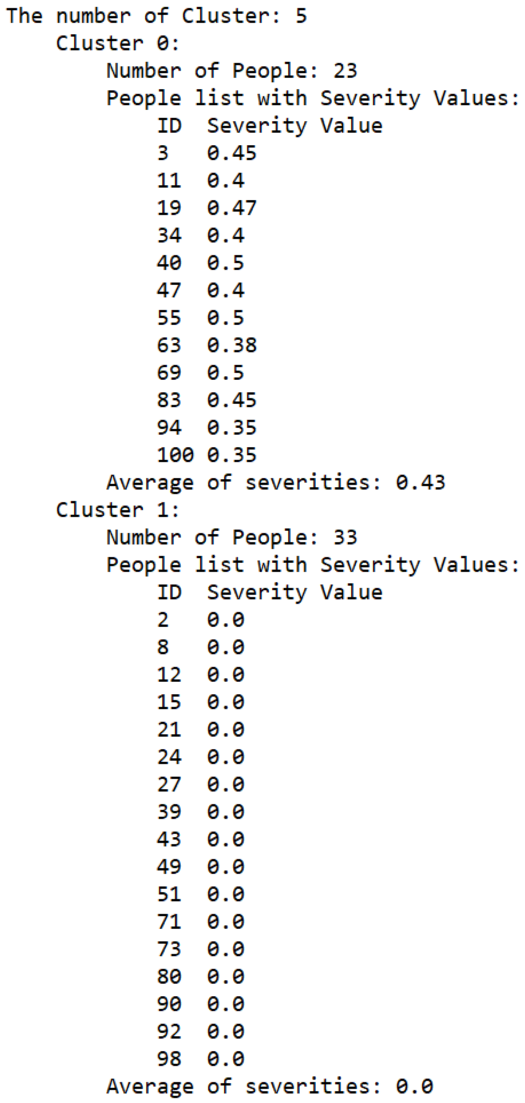
```
The number of Cluster: 4
    Cluster 0:
        Number of People: 33
        People list with Severity Values:
            ID  Severity Value
            2   0.0
            8   0.0
            12  0.0
            15  0.0
            21  0.0
            24  0.0
            27  0.0
            39  0.0
            43  0.0
            49  0.0
            51  0.0
            71  0.0
            73  0.0
            80  0.0
            90  0.0
            92  0.0
            98  0.0
        Average of severities: 0.0
```

- Cluster 1, 2, 3



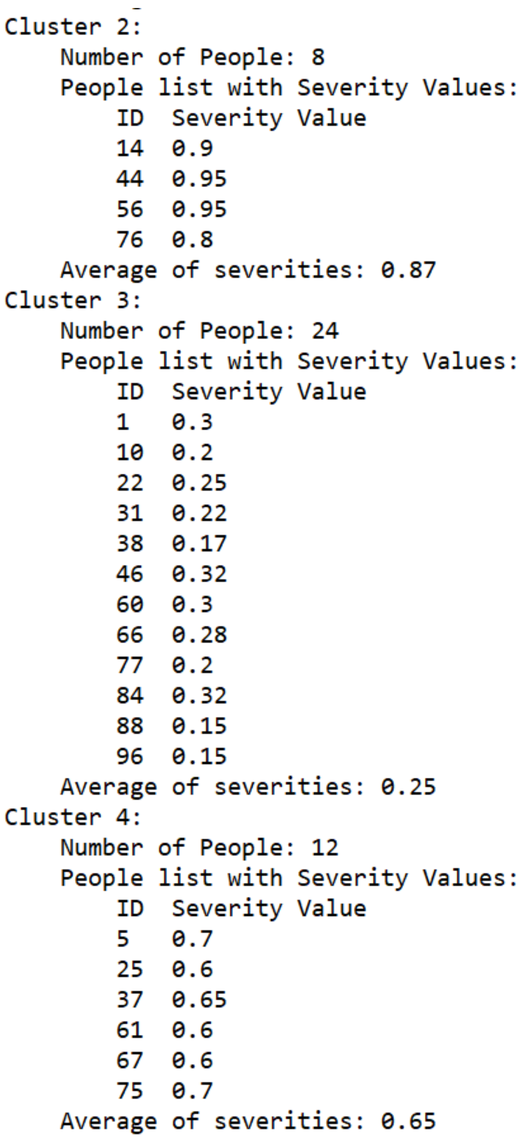
```
Cluster 1:
    Number of People: 30
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        10  0.2
        20  0.28
        29  0.25
        32  0.17
        42  0.3
        53  0.17
        62  0.25
        66  0.28
        77  0.2
        84  0.32
        87  0.3
        89  0.15
        96  0.15
        99  0.3
    Average of severities: 0.27
Cluster 2:
    Number of People: 15
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        18  0.65
        37  0.65
        52  0.8
        57  0.65
        74  0.7
        76  0.8
        95  0.7
    Average of severities: 0.78
Cluster 3:
    Number of People: 22
    People list with Severity Values:
        ID  Severity Value
        3   0.45
        11  0.4
        25  0.6
        30  0.6
        36  0.47
        45  0.4
        54  0.47
        59  0.45
        64  0.6
        67  0.6
        82  0.45
    Average of severities: 0.49
```

- **The number of Cluster: 5**
  - Cluster 0, 1



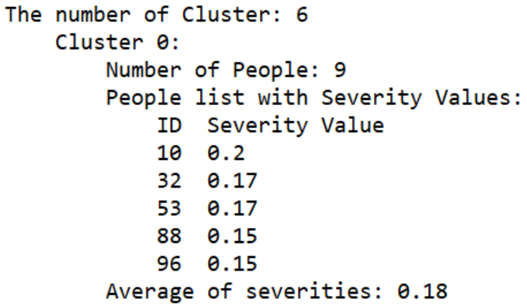
```
The number of Cluster: 5
    Cluster 0:
        Number of People: 23
        People list with Severity Values:
            ID  Severity Value
            3   0.45
            11  0.4
            19  0.47
            34  0.4
            40  0.5
            47  0.4
            55  0.5
            63  0.38
            69  0.5
            83  0.45
            94  0.35
            100 0.35
        Average of severities: 0.43
    Cluster 1:
        Number of People: 33
        People list with Severity Values:
            ID  Severity Value
            2   0.0
            8   0.0
            12  0.0
            15  0.0
            21  0.0
            24  0.0
            27  0.0
            39  0.0
            43  0.0
            49  0.0
            51  0.0
            71  0.0
            73  0.0
            80  0.0
            90  0.0
            92  0.0
            98  0.0
        Average of severities: 0.0
```

  - Cluster 2, 3, 4



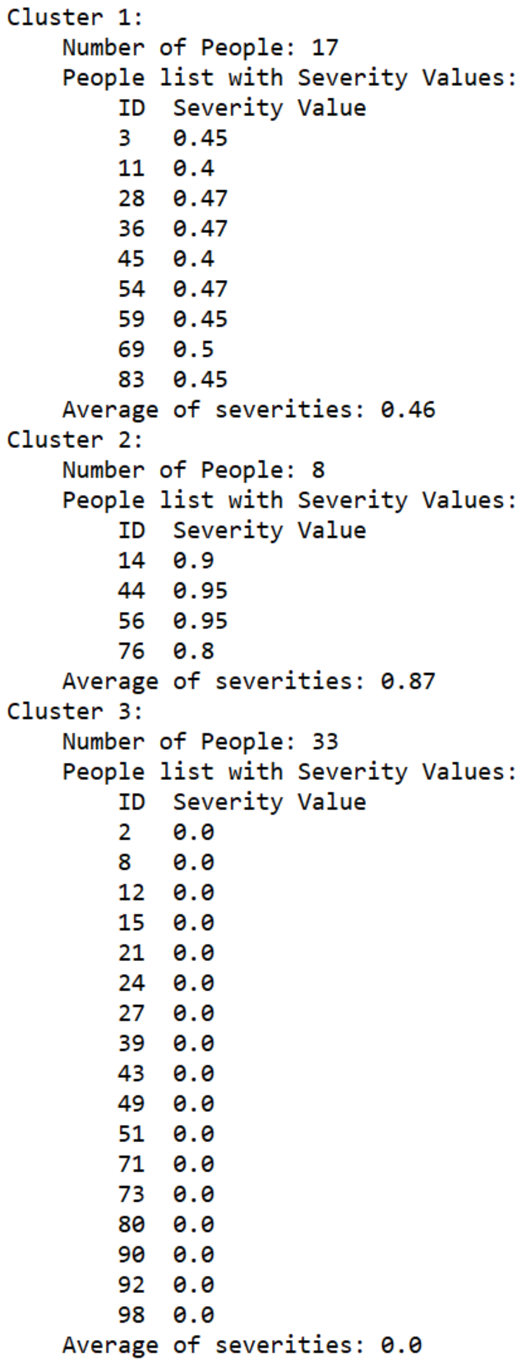
```
Cluster 2:
    Number of People: 8
    People list with Severity Values:
        ID  Severity Value
        14  0.9
        44  0.95
        56  0.95
        76  0.8
    Average of severities: 0.87
Cluster 3:
    Number of People: 24
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        10  0.2
        22  0.25
        31  0.22
        38  0.17
        46  0.32
        60  0.3
        66  0.28
        77  0.2
        84  0.32
        88  0.15
        96  0.15
    Average of severities: 0.25
Cluster 4:
    Number of People: 12
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        25  0.6
        37  0.65
        61  0.6
        67  0.6
        75  0.7
    Average of severities: 0.65
```

- ❑ **The number of Cluster: 6**
  - ❍ Cluster 0



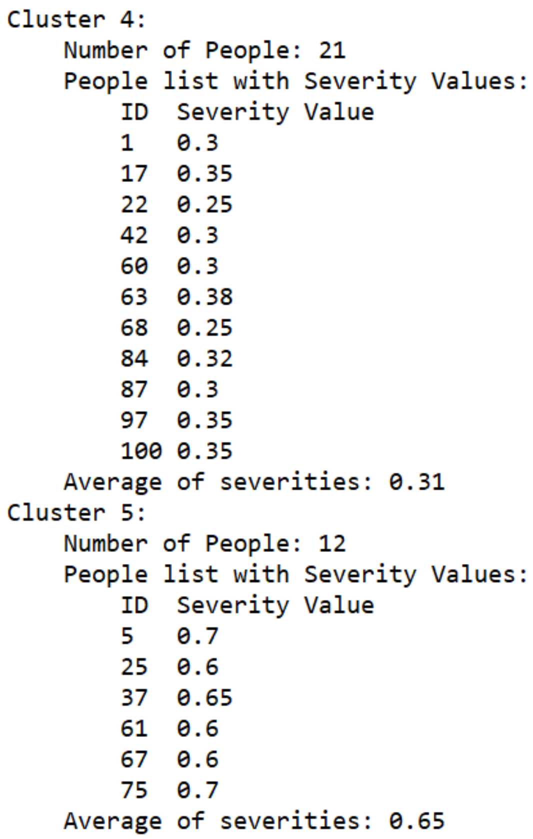
```
The number of Cluster: 6
    Cluster 0:
        Number of People: 9
        People list with Severity Values:
            ID  Severity Value
            10  0.2
            32  0.17
            53  0.17
            88  0.15
            96  0.15
        Average of severities: 0.18
```

  - ❍ Cluster 1, 2, 3



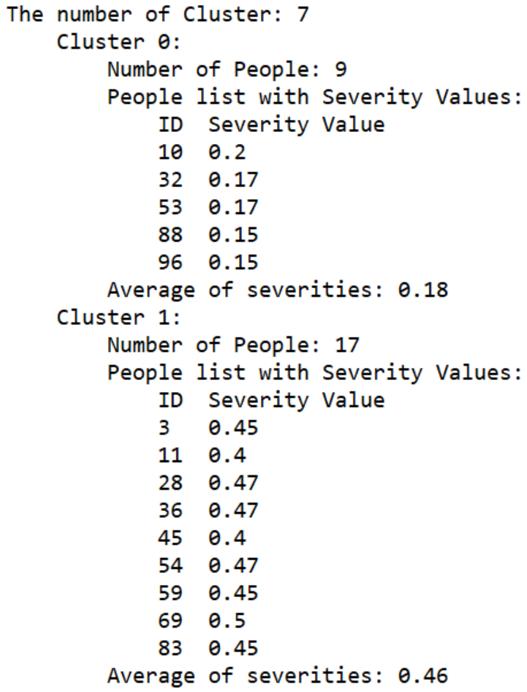
```
Cluster 1:
    Number of People: 17
    People list with Severity Values:
        ID  Severity Value
        3   0.45
        11  0.4
        28  0.47
        36  0.47
        45  0.4
        54  0.47
        59  0.45
        69  0.5
        83  0.45
    Average of severities: 0.46
Cluster 2:
    Number of People: 8
    People list with Severity Values:
        ID  Severity Value
        14  0.9
        44  0.95
        56  0.95
        76  0.8
    Average of severities: 0.87
Cluster 3:
    Number of People: 33
    People list with Severity Values:
        ID  Severity Value
        2   0.0
        8   0.0
        12  0.0
        15  0.0
        21  0.0
        24  0.0
        27  0.0
        39  0.0
        43  0.0
        49  0.0
        51  0.0
        71  0.0
        73  0.0
        80  0.0
        90  0.0
        92  0.0
        98  0.0
    Average of severities: 0.0
```

- Cluster 4, 5



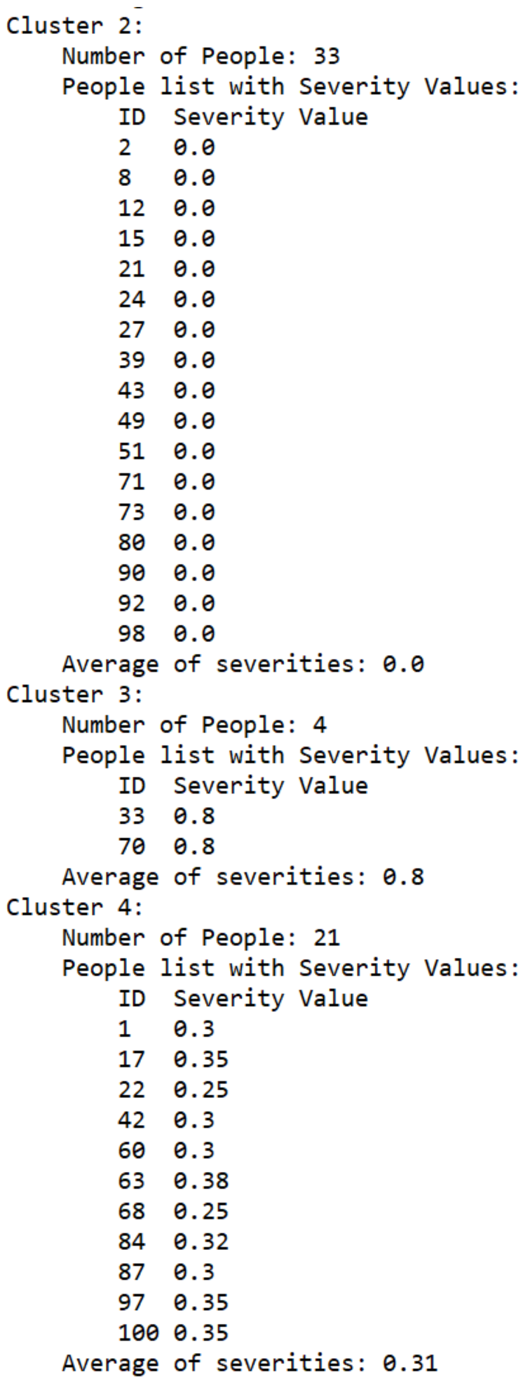
```
Cluster 4:
    Number of People: 21
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        17  0.35
        22  0.25
        42  0.3
        60  0.3
        63  0.38
        68  0.25
        84  0.32
        87  0.3
        97  0.35
        100 0.35
    Average of severities: 0.31
Cluster 5:
    Number of People: 12
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        25  0.6
        37  0.65
        61  0.6
        67  0.6
        75  0.7
    Average of severities: 0.65
```

- **The number of Cluster: 7**
  - Cluster 0, 1



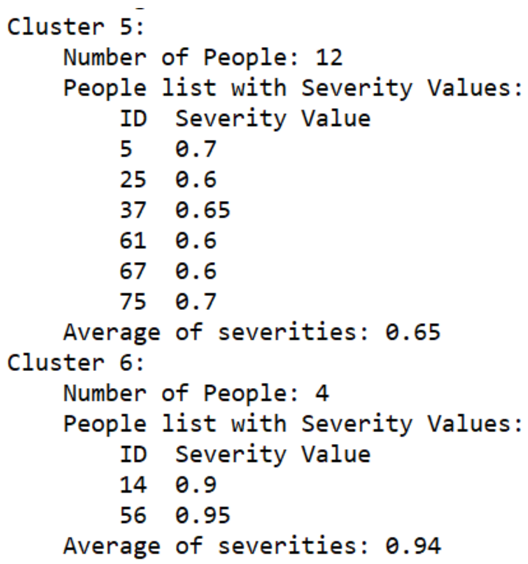
```
The number of Cluster: 7
    Cluster 0:
        Number of People: 9
        People list with Severity Values:
            ID  Severity Value
            10  0.2
            32  0.17
            53  0.17
            88  0.15
            96  0.15
        Average of severities: 0.18
    Cluster 1:
        Number of People: 17
        People list with Severity Values:
            ID  Severity Value
            3   0.45
            11  0.4
            28  0.47
            36  0.47
            45  0.4
            54  0.47
            59  0.45
            69  0.5
            83  0.45
        Average of severities: 0.46
```

  - Cluster 2, 3, 4



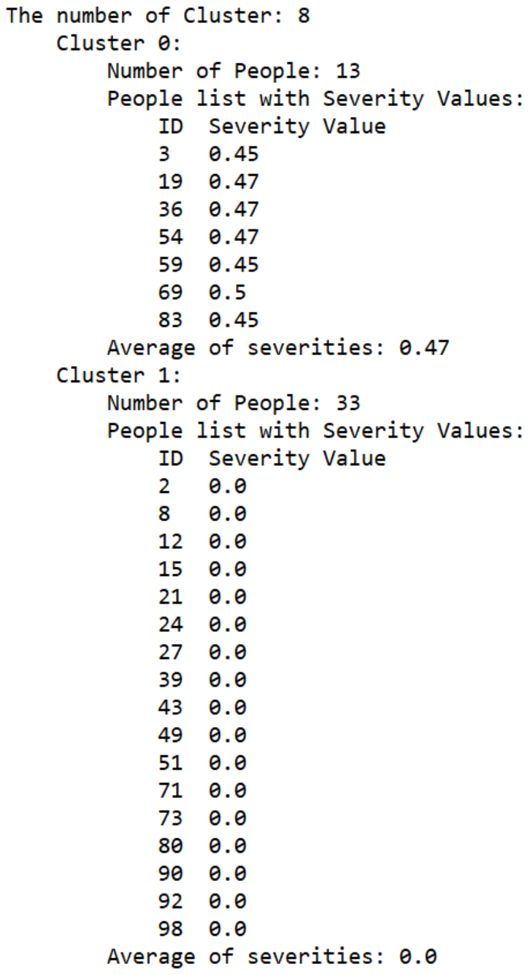
```
Cluster 2:
    Number of People: 33
    People list with Severity Values:
        ID  Severity Value
        2   0.0
        8   0.0
        12  0.0
        15  0.0
        21  0.0
        24  0.0
        27  0.0
        39  0.0
        43  0.0
        49  0.0
        51  0.0
        71  0.0
        73  0.0
        80  0.0
        90  0.0
        92  0.0
        98  0.0
    Average of severities: 0.0
Cluster 3:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        33  0.8
        70  0.8
    Average of severities: 0.8
Cluster 4:
    Number of People: 21
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        17  0.35
        22  0.25
        42  0.3
        60  0.3
        63  0.38
        68  0.25
        84  0.32
        87  0.3
        97  0.35
        100 0.35
    Average of severities: 0.31
```

❍ Cluster 5, 6



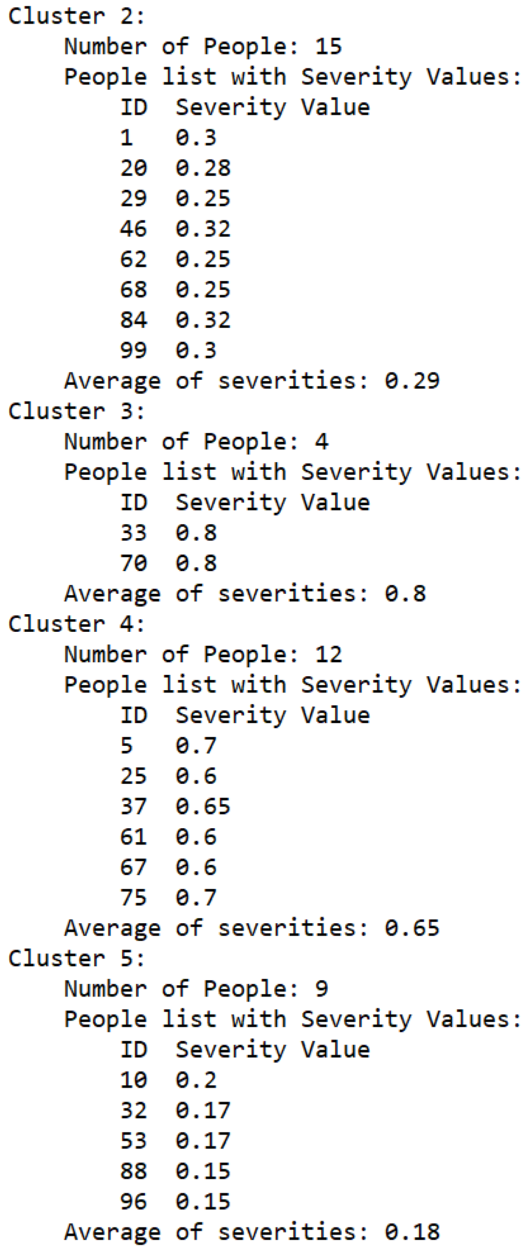
```
Cluster 5:
    Number of People: 12
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        25  0.6
        37  0.65
        61  0.6
        67  0.6
        75  0.7
    Average of severities: 0.65
Cluster 6:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        14  0.9
        56  0.95
    Average of severities: 0.94
```

- ❑ **The number of Cluster: 8**
  - ❍ Cluster 0, 1



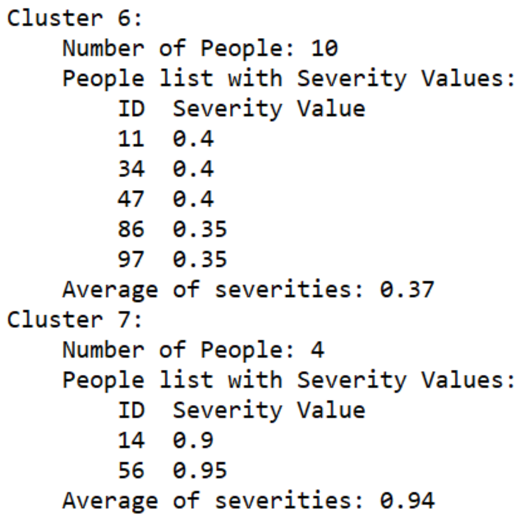
```
The number of Cluster: 8
    Cluster 0:
        Number of People: 13
        People list with Severity Values:
            ID  Severity Value
            3   0.45
            19  0.47
            36  0.47
            54  0.47
            59  0.45
            69  0.5
            83  0.45
        Average of severities: 0.47
    Cluster 1:
        Number of People: 33
        People list with Severity Values:
            ID  Severity Value
            2   0.0
            8   0.0
            12  0.0
            15  0.0
            21  0.0
            24  0.0
            27  0.0
            39  0.0
            43  0.0
            49  0.0
            51  0.0
            71  0.0
            73  0.0
            80  0.0
            90  0.0
            92  0.0
            98  0.0
        Average of severities: 0.0
```

  - ❍ Cluster 2, 3, 4, 5



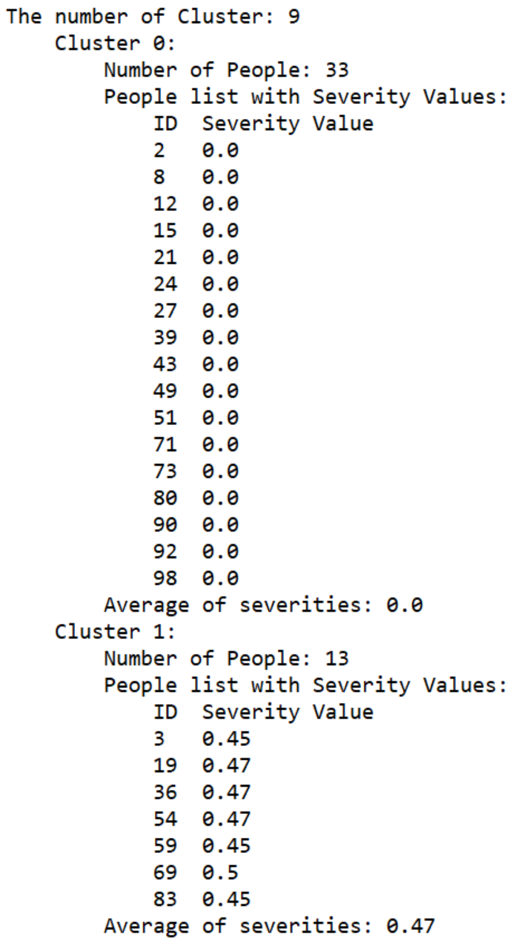
```
Cluster 2:
    Number of People: 15
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        20  0.28
        29  0.25
        46  0.32
        62  0.25
        68  0.25
        84  0.32
        99  0.3
    Average of severities: 0.29
Cluster 3:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        33  0.8
        70  0.8
    Average of severities: 0.8
Cluster 4:
    Number of People: 12
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        25  0.6
        37  0.65
        61  0.6
        67  0.6
        75  0.7
    Average of severities: 0.65
Cluster 5:
    Number of People: 9
    People list with Severity Values:
        ID  Severity Value
        10  0.2
        32  0.17
        53  0.17
        88  0.15
        96  0.15
    Average of severities: 0.18
```

- Cluster 6, 7



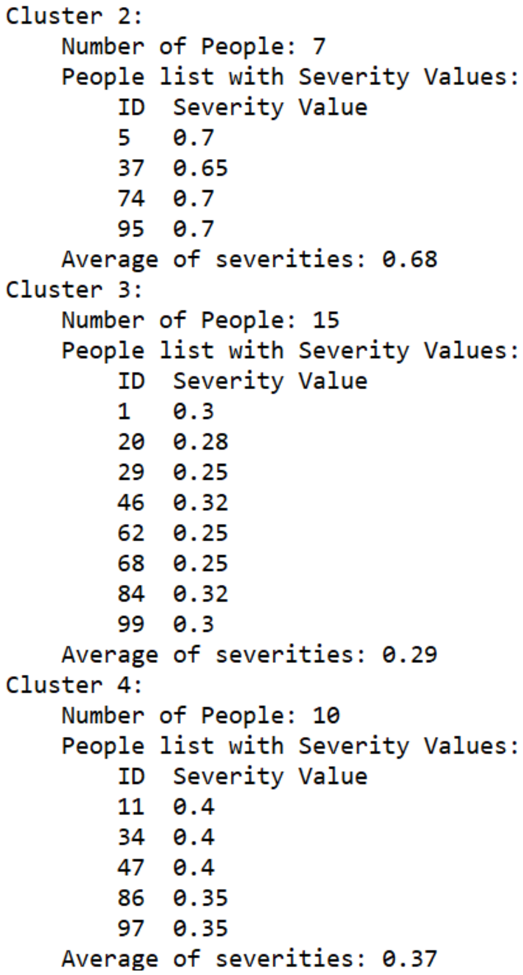
```
Cluster 6:
    Number of People: 10
    People list with Severity Values:
        ID  Severity Value
        11  0.4
        34  0.4
        47  0.4
        86  0.35
        97  0.35
    Average of severities: 0.37
Cluster 7:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        14  0.9
        56  0.95
    Average of severities: 0.94
```

- **The number of Cluster: 9**
  - Cluster 0, 1



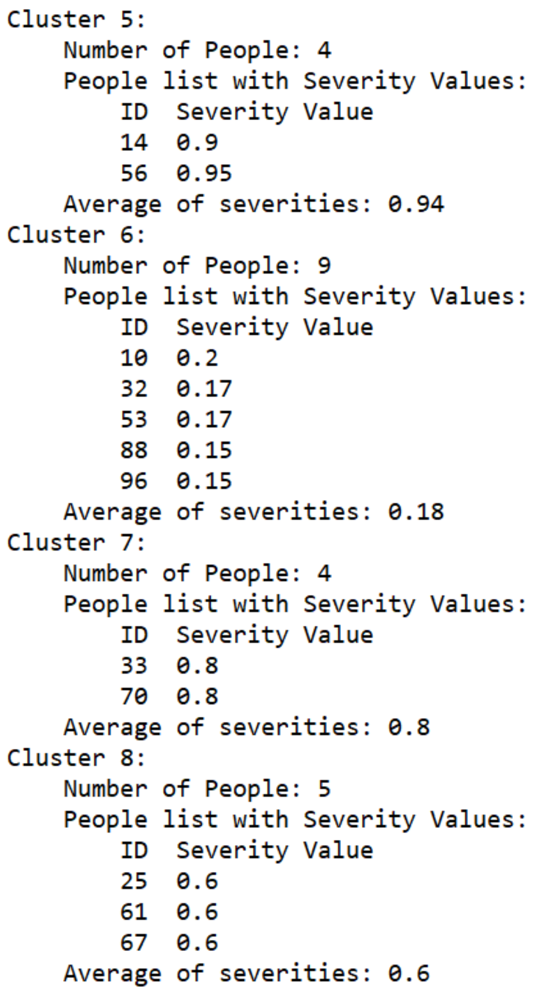
```
The number of Cluster: 9
    Cluster 0:
        Number of People: 33
        People list with Severity Values:
            ID  Severity Value
            2   0.0
            8   0.0
            12  0.0
            15  0.0
            21  0.0
            24  0.0
            27  0.0
            39  0.0
            43  0.0
            49  0.0
            51  0.0
            71  0.0
            73  0.0
            80  0.0
            90  0.0
            92  0.0
            98  0.0
        Average of severities: 0.0
    Cluster 1:
        Number of People: 13
        People list with Severity Values:
            ID  Severity Value
            3   0.45
            19  0.47
            36  0.47
            54  0.47
            59  0.45
            69  0.5
            83  0.45
        Average of severities: 0.47
```

  - Cluster 2, 3, 4



```
Cluster 2:
    Number of People: 7
    People list with Severity Values:
        ID  Severity Value
        5   0.7
        37  0.65
        74  0.7
        95  0.7
    Average of severities: 0.68
Cluster 3:
    Number of People: 15
    People list with Severity Values:
        ID  Severity Value
        1   0.3
        20  0.28
        29  0.25
        46  0.32
        62  0.25
        68  0.25
        84  0.32
        99  0.3
    Average of severities: 0.29
Cluster 4:
    Number of People: 10
    People list with Severity Values:
        ID  Severity Value
        11  0.4
        34  0.4
        47  0.4
        86  0.35
        97  0.35
    Average of severities: 0.37
```

- Cluster 5, 6, 7, 8



```
Cluster 5:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        14  0.9
        56  0.95
    Average of severities: 0.94
Cluster 6:
    Number of People: 9
    People list with Severity Values:
        ID  Severity Value
        10  0.2
        32  0.17
        53  0.17
        88  0.15
        96  0.15
    Average of severities: 0.18
Cluster 7:
    Number of People: 4
    People list with Severity Values:
        ID  Severity Value
        33  0.8
        70  0.8
    Average of severities: 0.8
Cluster 8:
    Number of People: 5
    People list with Severity Values:
        ID  Severity Value
        25  0.6
        61  0.6
        67  0.6
    Average of severities: 0.6
```